

**STIC-ILL**

NOT owned

**From:** Goldberg, Jeanine  
**Sent:** Friday, April 20, 2001 3:46 PM  
**To:** STIC-ILL  
**Subject:** pleaesa pul multiplex references

342744

1. PCR METHODS AND APPLICATIONS, (1993 Aug) 3 (1) 13-22.  
Journal code: BNV. ISSN: 1054-9803.

2. INTERNATIONAL JOURNAL OF LEGAL MEDICINE, (1994) 106 (6)  
302-11.  
Journal c

2922485

3. INTERNATIONAL JOURNAL OF LEGAL MEDICINE, (1994) 107 (1)  
13-20.  
Journal

4. FORENSIC SCIENCE INTERNATIONAL, (1997 Oct 6) 89 (3) 185-97.  
Journal code: F49. ISSN: 0379-0738.

THANKS

Jeanine Enewold Goldberg  
1655  
CM1--12D11  
Mailbox-- 12E12  
306-5817

2999641

## ORIGINAL ARTICLE

A. Urquhart · C. P. Kimpton · T. J. Downes · P. Gill

## Variation in Short Tandem Repeat sequences – a survey of twelve microsatellite loci for use as forensic identification markers

Received: 1 February 1994 / Received in revised form: 28 April 1994

**Abstract** Alleles at 12 Short Tandem Repeat loci have been sequenced to investigate candidate loci for a multiplex Short Tandem Repeat system for forensic identification, and for single-locus amplification of Short Tandem Repeat loci. Variation from the consensus sequence was found at 6 loci, while one locus, D21S11, was found to be complex in sequence. The presence of non-consensus alleles does not rule out loci for inclusion as forensic identification markers, but size differences between alleles of 1 base pair require very precise sizing. We suggest criteria for the suitability of Short Tandem Repeat loci as forensic identification markers, and propose a universal allele nomenclature for simple and compound Short Tandem Repeats. The effect of the repeat unit sequence of the evolution of Short Tandem Repeats is discussed.

**Key words** Short Tandem Repeats · Microsatellites  
DNA sequencing · Polymerase Chain Reaction · Forensic DNA typing

**Zusammenfassung** Allele an 12 Short-Tandem-Repeat Loci wurden sequenziert, um Kandidaten für ein Multiplex Short Tandem Repeat System für forensische Identifikationen und für Single-Locus Amplifikationen von Short-Tandem-Repeat Loci zu untersuchen. Abweichungen von der Konsensus-Sequenz wurden an 6 Loci gefunden, während ein Locus, D21S11, als Komplex in der Sequenz gefunden wurde. Die Anwesenheit von Non-Konsensus-Allelen schließt solche Loci nicht aus für die Einbeziehung als forensische Identifikationsmarker. Aber Größendifferenzen von einem Basenpaar zwischen Allelen erfordern eine sehr genaue Größenbestimmung. Wir empfehlen Kriterien für die Eignung von Short-Tandem-Repeat Loci als forensische Identifikationsmarker und schlagen eine universale Allelnomenklatur für einfache und kom-

plexe Short-Tandem-Repeats vor. Die Auswirkung der Sequenz der Repeateinheit auf die Entwicklung von Short-Tandem-Repeats wird diskutiert.

**Schlüsselwörter** Short-Tandem-Repeats  
Mikrosatelliten · DNA-Sequenzierung · Polymerase Kettenreaktion · Forensische DNA-Typisierung

### Introduction

Analysis of Short Tandem Repeat (STR) sequences by the polymerase chain reaction (PCR) is becoming the method of choice for the forensic identification of body fluids (Kimpton et al. 1993, 1994; Fregeau and Fournay 1993; Wiegand et al. 1993). Because of problems caused by 'shadow bands' when analysing dinucleotide repeats (Hauge and Litt 1993), the less common tri-, tetra- and pentanucleotide repeats are preferred.

STR sequences vary in the length of repeat unit, the number of repeats and the rigour with which they conform to an incremental repeat pattern. 'Simple' repeats contain units of identical length and sequence, 'compound' repeats comprise 2 or more adjacent simple repeats, 'complex' repeats may contain several repeat blocks of variable unit length, along with more or less variable intervening sequences.

We have recently studied sequence variation at 2 complex STR loci, HUMACTBP2 (SE33) and D11S554 (Urquhart et al. 1993; Adams et al. 1993). Both these loci were originally reported to have allele sizes which differed by 4 base pair increments (Warne et al. 1991; Phromchotikul et al. 1992). However, our sequence data showed that at both loci allele size differences of 1, 2 or 3 base pairs also exist.

Since allele designation of STR PCR products depends on accurate sizing, we investigated a range of simple, compound and complex STR loci which were being screened in this laboratory for use as forensic identification markers. The markers used in our quadruplex STR system (Kimpton et al. 1993, 1994), a *Pst*I digest of bacteriophage lambda

A. Urquhart (✉) · C. P. Kimpton · T. J. Downes · P. Gill  
Central Research and Support Establishment,  
The Forensic Science Service, Birmingham B5 6QQ, UK

Table 1 PCR primers used

Locus	Primers	Reference
HUMVWFA31	5' CCCTAGTGGATGATAAGAATAATCAGTATG 3' GGACAGATGATAAATACATAGGATGGATGG	Kimpton et al. 1992, GenBank M25858
HUMTH01	5' GTGATTCCCATTGGCCTGTTCTCCTC 3' GTGGGCTGAAAAGCTCCCGATTAT	Polymeropoulos et al. 1991f, GenBank D00269
HUMF13A01	5' GAGGTTGCACTCCAGCCTTT 3' ATGCCATGCAGATTAGAAA	Polymeropoulos et al. 1991cm, GenBank M21986
HUMFES/FPS	5' GGGATTTCCCTATGGATTGG 3' GCGAAAGAATGAGACTACAT	Polymeropoulos et al. 1991b, GenBank X06292
HUMCD4	5' TTGGAGTCGCAAGCTGAACTAGC 3' TCATGCGTCCATGGTCCGGAGCCTGAGTGACAGAGTGAGAACC	Edwards et al. 1991, GenBank M86525
HUMPLA2A1	5' CCCACTAGGTTGTAAGCTCCATGA 3' TACTATGTGCCAGGCTCTGTCTTA	Polymeropoulos et al. 1990b, GenBank M22970
HUMFOLP23	5' ATTGTAAGACTTTTGGAGCCATTT 3' TTCAGGGAGAATGAGATGGGC	Polymeropoulos et al. 1991d, GenBank J00145
HUMCYAR04	5' CTCTGGAACCAACTCGACCCTTC 3' TGGGTGATAGAGTCAGAGCCTGTC	Polymeropoulos et al. 1991a, GenBank M30798
HUMTFIIDA	5' GCCTATTCAGAACACCAATA 3' TGGGACGTTGACTGCTGAAC	Polymeropoulos et al. 1991e, GenBank M34960
HUMFABP	5' GTAGTATCAGTTTCATAGGGTCACC 3' TTACGCGTCTCGACAGTATTTCAGTTCGTTTC	Polymeropoulos et al. 1990a, GenBank M18079
HUMGABRB15	5' CTAGAAAGCTAGCAAGGTGGAT 3' GCTCATTAAACACTGTGTTCTCCT	Dean et al. 1991, GenBank M59216
HUMD21S11	5' ATATGTGAGTCAATTCCTCAAG 3' TGTATTAGTCAATGTTCTCCAG	Sharma and Litt 1992, GenBank M84567

DNA labelled with the fluorescent dye ROX, sized alleles precisely but not accurately, and in our hands sizing alleles differing by only 1 bp could not be performed without the use of an allelic ladder. Since many laboratories are now involved in STR analysis, and before STR data becomes widespread as courtroom evidence, it would be convenient if a universal system of allele designation and nomenclature were adopted. To this end, we have investigated 12 prospective human STR identification markers to ascertain convenient, easily understood and scientifically accurate methods of allele nomenclature.

### Materials and methods

The 12 loci studied and their respective PCR primer sequences are shown in Table 1. All primers were derived from the published or GenBank sequences, but the HUMFABP 3' primer was lengthened to incorporate a *MluI* restriction site, and the HUMCD4 5' primer had a 20 bp extension, designed by Jeffreys and co-workers (Jeffreys et al. 1991), to produce allele sizes compatible with one of our multiplex STR systems (Kimpton et al. 1993). DNA was prepared from whole blood as described previously (Gill et al. 1990). Allele designations at each locus had been made previously (Kimpton et al. 1993). For sequencing, heterozygotes with allele sizes differing by at least 12, and ideally 20, base pairs, were selected, and between 8 and 22 alleles were sequenced at each locus. This was intended to be a representative sample, rather than an exhaustive survey, of alleles at each locus.

PCR amplification was performed using 10 ng of genomic DNA in a 50 µl reaction volume. Reactions included 1 × Parr-Excellence buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1% Triton X-100; Cambio Labs, Cambridge, UK), 1.25 U Taq polymerase (Perkin Elmer Cetus, Norwalk, CT, USA), 200 µM each deoxyribonucleotide triphosphate and 0.5 µM each of 2 primers for each locus: 35 cycles of PCR (1 min 95°C, 1 min 54°C, 1 min 72°C) were performed for the loci HUMVWFA31, HUMF13A01 and HUMFES/FPS. For the other 9 loci the annealing temperature was 60°C; other conditions were identical and 35 cycles were again performed.

PCR products were electrophoresed in agarose gels, excised and purified as described previously (Urquhart 1991). Purified PCR products were sequenced from both ends with a Taq Dye-Deoxy Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, Calif., USA) using the PCR primers as sequencing primers. Sequence analysis was performed on a 373A Sequencer (Applied Biosystems, Foster City, Calif., USA) using 373 Data Collection, 373 Analysis and SeqEd software (Applied Biosystems, Foster City, Calif., USA).

### Results and discussion

The consensus sequences of repeat regions at the 12 loci are shown in Figs. 1–3. The repeat unit at each locus was defined as the first in-frame repeat unit on the strand listed in the GenBank database. Where necessary, ambiguity codes were used, in accordance with the recommendations of the Nomenclature Committee of the International Union

**LOCUS****Simple**

HUMFES/FPS	(ATTT) <sub>8-14</sub>
HUMPLA2A1	(ATT) <sub>8-17</sub>
HUMFABP	(ATT) <sub>8-15</sub>

**Simple with non-consensus alleles**

HUMTH01	(TCAT) <sub>5-11</sub>
173bp allele	(TCAT) <sub>4</sub> <u>CAT</u> (TCAT) <sub>5</sub>
HUMFOLP23	(AAAC) <sub>4-8</sub> AAAA.AAAC
174bp allele	(AAAC) <sub>10</sub> -----
HUMCD4	(TTTTC) <sub>3-13</sub>
141bp allele	(TTTTC) <sub>3</sub> <u>CTTTC</u> (TTTTC) <sub>8</sub>
HUMF13A01	(GAAA) <sub>4-17</sub> GAGTAAAA
181bp allele	(GAAA) <sub>5</sub> -----AA
HUMCYAR04	AT(CTT) <sub>2</sub> TTTTGTCTATGAATGTCCTTTTTTGAATCATATTTTAAATAT(TTTA) <sub>7-13</sub>
166bp allele	AT(CTT) <sub>1</sub> TTTTGTCTATGAATGTCCTTTTTTGAATCATATTTTAAATAT(TTTA) <sub>7</sub>

**Fig.1** Simple repeat sequences. The variable repeat regions are shown along with their flanking sequence where relevant. Differences from the consensus sequence for each locus are underlined

of Biochemistry (1985). Hence, M signifies A or C, Y signifies C or T, K signifies G or T, R signifies A or G and V signifies A, C or G.

Of the 12, 8 loci (HUMFES/FPS, HUMPLA2A1, HUMFABP, HUMTH01, HUMF13A01, HUMCYAR04, HUMFOLP23 [formerly HUMDHFRP2] and HUMCD4) were classified as simple repeats (Fig. 1), although HUMTH01, HUMF13A01 and HUMCYAR04 each had one common allele which differed from the consensus, while in HUMFOLP23 and HUMCD4 there was variation in individual repeats units. For HUMF13A01 and HUMCYAR04 the non-consensus allele was the smallest allele found, and each involved a deletion outside the repeat region (Fig. 1). The deletion in HUMCYAR04 is 1 of 2 CTT trinucleotides 51 bp 5' to the repeat region, while that in HUMF13A01 was a GAGTAA hexanucleotide immediately 3' to the repeat. This deletion occurred in all 6 of the 181 bp alleles sequenced. At the HUMTH01 locus, the largest common alleles, although a 178 bp allele is found very occasionally (Edwards et al. 1992; our unpublished

data), are sized at 173–174 bp. We have sequenced 11 alleles sized at 173/174 bp, of which 8 were 173 bp and 3 were 174 bp. All the 173 bp alleles had the same single-base deletion of a thymidine residue in the fifth of 10 TCAT repeats. These observations have recently been reported elsewhere (Puers et al. 1993).

The HUMFOLP23 locus has been called a simple repeat in this study, although most alleles contain the octamer AAAA.AAAC following the run of AAAC repeats. Nine alleles (including one 174 bp allele, the largest allele size found at this locus) had 4–8 AAAC repeats followed by AAAA.AAAC. One 174 bp allele had 10 AAAC repeats without the following 8 bp. Since the two 174 bp alleles are indistinguishable by band-sizing methods, we decided to designate both as alleles with 10 AAAM repeats. It is not known whether smaller alleles consisting solely of AAAC repeats exist. Similarly, the HUMCD4 locus has been called a simple repeat although both the GenBank sequence (Edwards et al. GenBank M86525) and one allele (out of 11) that we sequenced had CTTTC as the fourth repeat instead of the consensus TTTTC. Alleles were designated as YTTTC repeats.

Three loci (HUMGABRB15, HUMTFIIDA and HUMVWFA31) were classified as compound repeats (Fig. 2).

**Fig.2** Compound repeat sequences. The variable repeat regions are shown. Differences from the consensus sequence for the HUMVWFA31 locus are underlined

**LOCUS****Compound**

HUMGABRB15	(GATA) <sub>5-12</sub> (GATC) <sub>2-4</sub> (TATC) <sub>1-2</sub>
HUMTFIIDA	(CAG) <sub>3</sub> (CAA) <sub>3</sub> (CAG) <sub>9-11</sub> CAA(CAG.CAA) <sub>0-1</sub> (CAG) <sub>9-24</sub> CAA.CAG

**Compound with non-consensus alleles**

HUMVWFA31	(ATCT) <sub>2</sub> (GTCT) <sub>3-4</sub> (ATCT) <sub>9-13</sub>
144bp allele	(ATCT) <sub>2</sub> (GTCT) <sub>4</sub> (ATCT) <sub>5</sub> <u>AT</u> (ATCT) <sub>4</sub>

	x	y
GenBank	(TCTA) <sub>4</sub> (TCTG) <sub>6</sub> (TCTA) <sub>3</sub> TA(TCTA) <sub>3</sub> TCA(TCTA) <sub>2</sub> ----- (TCTA) <sub>8</sub> -----TC	
Odd	(TCTA) <sub>4-6</sub> (TCTG) <sub>5-6</sub> (TCTA) <sub>3</sub> TA(TCTA) <sub>3</sub> TCA(TCTA) <sub>2</sub> <u>TCCATA</u> (TCTA) <sub>8-16</sub> -----TC	
Even	(TCTA) <sub>4-6</sub> (TCTG) <sub>5-6</sub> (TCTA) <sub>3</sub> TA(TCTA) <sub>3</sub> TCA(TCTA) <sub>2</sub> <u>TCCATA</u> (TCTA) <sub>8-16</sub> <u>TA.TCTA.TC</u>	
GenBank	GTCTATCTATCCAGTCTATCTACCTCTATNNAG	
Odd	GTCTATCTATCCAGTCTATCTACCTCTAT <u>T</u> -AG	
Even	GTCTATCTATCCAGTCTATCTACCTCTAT <u>T</u> -AG	

Fig. 3 Complex repeat sequence at the D21S11 locus. The variable repeat region is shown along with 33 bp of 3' sequence for the GenBank sequence (Sharma and Litt 1992) and our consensus sequences for the odd-numbered and even-numbered alleles. The invariant hexanucleotide which we found in all alleles and the 2 amendments we made to the GenBank sequence are singly underlined. The hexanucleotide found in all even-numbered alleles is doubly underlined. The segments marked x and y are those used for allele designation (see text)

The HUMGABRB15 locus consists of blocks of GATA, GATC and TATC repeats, all of which vary in number between alleles. The aggregate number of the 3 repeat types (i.e. the number of KATM repeats) was used for allele designation. Similarly, the HUMTFIIDA locus contains 7 or 9 blocks of CAG or CAA repeats, and allele designation was for CAR.

Apart from one allele in one individual, HUMVWFA31 was a straightforward compound repeat with the sequence (ATCT)<sub>2</sub>(GTCT)<sub>m</sub>(ATCT)<sub>n</sub>, and allele designation is for RTCT. However, one non-consensus allele of 144 bp was observed in which the 3'ATCT tract contained an AT dinucleotide (Fig. 2). This could have arisen either by deletion of TC from a 146 bp 16 allele or by duplication of TA in a 142 bp 15 allele. This allele was the only HUMVWFA31 allele seen which differed from the 4 bp repeat pattern in over 1500 alleles sized at the locus (unpublished data).

The repeat at the D21S11 locus was classified as a complex repeat. Although originally reported to have 4 bp increments between alleles (Sharma and Litt 1992), later work (Kimpton et al. 1993) revealed that alleles differing by 2 bp were common. The consensus sequences of the repeat at the D21S11 locus are shown in Fig. 3, aligned with the sequence from GenBank (Accession number M84567; Sharma and Litt 1992). A total of 16 alleles was sequenced, including the largest and smallest, and 2 pairs of identically sized alleles, one pair of which had identical sequence. The variable sequence consisted largely of TCTA and TCTG repeat blocks, although an invariant TA dinucleotide and an invariant TCA trinucleotide were also present. Alleles which were given even-numbered allele designations (see 'Allele designation and nomenclature' below) had a TATCTA hexanucleotide after the final block of TCTA repeats. Alternatively, this can be viewed as a TA insertion before the last TCTA repeat. In all 16 alleles sequenced, a TCCATA hexanucleotide was found which does not appear in the GenBank sequence (Sharma and Litt 1992). If the absence of this hexanucleotide genuinely occurs, this is a further mechanism for 2 bp allele differences. Sequencing the D21S11 locus also allowed us to make 2 amendments to the GenBank sequence. In all 16

alleles sequenced, the N at position 288 in the GenBank sequence was a C, and the NN at positions 295-6 was a single T residue (Fig. 3).

#### Allele designation and nomenclature

Designation and nomenclature of alleles at STR and VNTR loci has been fairly haphazard in the past. For presentation of forensic evidence in the courts, and for transfer of data between different laboratories, some standardisation of allele designation is necessary. The most widely applicable method would be to call each allele by its length in basepairs. This method would be suitable for VNTRs, normal STRs and hypervariable STRs such as HUMACTBP2 (Urquhart et al. 1993) and D11S554 (Adams et al. 1993). However, the allele size is dependent on the primers used, and requires a precise and accurate sizing method. An alternative is to call alleles by the number of repeat units they contain. This is easy for simple repeats and some VNTRs, and can be applied to compound repeats with the use of ambiguity codes, but is too cumbersome for complex repeats. Problems also arise when intermediate alleles occur, as with the HUMVWFA31 144 bp allele in this study and the various anodic and cathodic allele variants in some VNTR systems. Both allele designation methods discussed above may involve loss of informativeness, since the repeat pattern at any individual allele is not specified. However, this is inevitable with all methods which distinguish alleles solely by size, and the increase in informativeness gained by sequencing every allele is more than offset by the increased cost. In line with the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics (1992), we have called alleles at all simple and compound repeat loci by their repeat number, using redundancy codes for compound repeats. For intermediate alleles and other alleles that fail to align with the incremental 'ladder' at each locus, digits after a decimal point were used to indicate the number of basepairs by which the allele exceeded the previous 'rung' of the ladder. Thus, the 144 bp allele at the HUMVWFA31 locus was designated 15.2, the 166 bp allele at the HUMCYAR04 locus was designated 6.1, the 173 bp allele at the HUMTH01 locus was designated 9.3, and the 181 bp allele at the HUMF13A01 locus was designated 3.2. It should be noted that the use of the number after the decimal point does not necessarily imply the presence of a partial repeat (cf. HUMF13A01, HUMCYAR04), but many indicate variation outside the repeat region.

**Table 2** Characteristics of (TCTA)<sub>5</sub>-containing human loci in GenBank

Locus	Accession number	Reference	Sequence motif			
			Alternative repeat (longest block)		Shortened repeat <sup>a</sup> (number present)	
			TCTG	TCCA	TCA	TA
D21S11	M84567	This study	5-6	1	1	2/3
HUMVWFA31(A) <sup>b</sup>	M25858	This study	3-4	2	-	0/1
HUMVWFA31(B) <sup>b</sup>	"	Mancuso et al. 1989	2	-	1	-
HUMVWFA31(C) <sup>b</sup>	"	Mancuso et al. 1989	1	-	2	-
HUMGABRB15	M59216	This study	-	-	-	-
T03428	T03428	Khan et al. 1992	-	-	-	-
D18S53	Z16461	Weissenbach et al. 1992	-	1	-	-
D2S121	Z16545	Weissenbach et al. 1992	-	-	-	-
D7S513	Z16989	Weissenbach et al. 1992	1	-	-	-
D10S225	Z17156	Weissenbach et al. 1992	1	2	1	1
HS7SKP41	X04237	Murphy et al. 1984	1	-	-	-
HSCSF1PO	X14720	Hampe et al. 1989	-	-	-	-
HSIGK6	Z00004	Jaenichen et al. 1984	-	-	-	-
HUMBKM	M35828	Erickson et al. 1988	1	-	-	2
HUMCD4(A) <sup>b</sup>	M86525	Edwards et al. GenBank	-	-	1	-
HUMCD4(B) <sup>b</sup>	"	Edwards et al. GenBank	1	-	-	1
HUMCD4(D) <sup>b</sup>	"	Edwards et al. GenBank	2	2	1	-
HUMHPRTB	M26434	Ansorge et al. 1990	1	-	-	-
HUMRPTPOLF	L02066	Weber and May 1989	-	-	1	-
HUMSIRPOBD	M87698	Hudson et al. 1992	-	-	-	-
HUMSIRPOCP	M87736	Hudson et al. 1992	1	-	1	-
HUMPGK1	S75476	Riley et al. 1991	1	-	-	-
HUMPAH	L10305	Goltsov et al. 1993	1	-	3	-
DXS981	M38419	Mahtani and Willard 1993	1	-	0/1	-
D12S66	-	Roewer et al. 1992	1	?	-	-
D12S67	-	Roewer et al. 1992	6	?	-	-
DYS19	-	Roewer et al. 1992	1	?	-	1

<sup>a</sup> TCA trinucleotides and TA dinucleotides were only included if they had at least one TCTA, TCTG or TCCA tetranucleotide repeat immediately adjacent on each side

<sup>b</sup> HUMVWFA31 A region: bases 1683-1770; B region: bases 1889-2063; C region: bases 2084-2343. HUMCD4 A region: bases 5624-5686; B region: bases 5944-6043; D region: bases 7101-7340

<sup>c</sup> Full sequence not given in reference

Allele designation for the complex repeat at D21S11 is more problematical. Each allele contains a mixture of di-, tri-, tetra- and hexanucleotides (Fig. 3). Three options were considered: naming alleles by length in base pairs, using an arbitrary system of allele designation, and naming alleles by the number of TV dinucleotide repeats. We decided on the third option largely because it was consistent with nomenclature at the other loci in this system. However, there were 2 minor inconsistencies. The system of nomenclature excludes the invariant TCA trinucleotide, and treats the CA in the centre of the TCCATA hexanucleotide as a TV dinucleotide. The allele designation at the D21S11 locus thus indicates the aggregate number of TV dinucleotides (plus one CA dinucleotide) in the two regions labelled x and y in Fig. 3.

### TCTA and related repeats

We noted several similarities between the sequences of D21S11 and HUMVWFA31. When written as TCTA repeats, both contain compound (TCTG)<sub>p</sub>(TCTA)<sub>q</sub> regions and the sequence motif (TCTA)<sub>r</sub>TA(TCTA)<sub>s</sub> appears (either once or twice) in D21S11 alleles and in the 144 bp non-consensus HUMVWFA31 allele. A search of GenBank human sequences for (TCTA)<sub>5</sub> and its complement (TAGA)<sub>5</sub> produced 22 matches, including D21S11, HUMGABRB15 and 3 regions of HUMVWFA31 (Table 2). These sequences, plus 5 others recently published, DXS981 (Mahtani and Willard 1993), phenylalanine hydroxylase (Goltsov et al. 1993), D12S66, D12S67 and DYS19 (Roewer et al. 1992), were examined for sequence motifs common to several sequences. These are summarised in Table 2. Many of the TCTA repeat blocks had single repeat units that differed by one base from the consensus repeat, the commonest being TCTG and TCCA

(occurring in 67% and 21% of sequences respectively). At some loci, notably D21S11, HUMVWFA31 and D12S67, longer blocks of TCTG were present. Truncated repeats were also present at some loci. The commonest trinucleotide, TCA, occurred at 37% of the loci, while the dinucleotide TA was present in 22% of loci. The frequency of these truncated repeats may well be an underestimate, since many of the reported TCTA repeats were discovered by hybridisation to  $(TAGA)_n$  or similar oligonucleotides, the binding of which would be decreased by imperfect repeats.

Searches were also performed for  $(TCTG)_5$  and  $(TCCA)_5$ , but both failed to find sequences associated with extensive TCTA repeats. However, the TCCA repeat in the HUMIGCAA locus (Yu et al. 1990) included 2 separate TCA trinucleotides. Interestingly, the non-consensus 9.3 allele at the HUMTH01 locus (see above) can be regarded as a TCA trinucleotide in the middle of a  $(TTCA)_n$  tract.

The sequence at the HUMGABRB15 locus (Fig. 2) can be written as  $(TAGA)_{5-12}(TCGA)_{1-3}(TCTA)_{1-2}$ , i.e. 2 mutually palindromic TCTA tracts surrounding a TCGA tract. It is possible that the central tract is formed by limited gene conversion, the 2 TCTA tracts in opposite orientation acting on each other.

These observations suggest a scenario for evolution of TCTA repeats. Since the TCA trinucleotide could not be generated by duplication from  $(TCTA)_n$ , we would suggest that this unit originates by deletion of a thymidine residue from TCTA. By analogy, TA dinucleotides may arise by deletion of C from TCA or, alternatively, by deletion of TC or CT from TCTA. Of course, they could also arise by TA duplication. In a simple TCTA repeat (e.g. D2S121) individual TCTA units may mutate to TCTG (or TCCA), giving imperfect but essentially simple repeats (e.g. D7S513). These imperfections may then expand, either by genuine size expansion or by gene conversion (Jackson and Fink 1981; Slightom et al. 1988), producing a compound repeat such as HUMVWFA31. TCTA (and TCCA) units may also undergo deletion to TCA or TA, producing complex repeats such as D21S11. Variation at the D21S11 locus suggests that repeat expansion can continue after event such as TCA generation. These events would lead to degeneration of simple TCTA repeats with time into complex repeats, containing TCTA, TCTG and TCCA blocks interspersed with dinucleotide and trinucleotide truncated repeats. Indeed, some degenerate TCTA repeats such as those at the HUMVWFA31 and HUMCD4 loci (Mancuso et al. 1989; Edwards et al. Gen Bank M86525) can reach thousands of bp in length.

#### Other repeats

As discussed above, the most common mutation in TCTA repeats is to TCTG, i.e. and A > G transition. The HUMTH01 and HUMCD4 loci appear to have developed by similar events, respectively by CAG > CAA and AAAAG > GAAAG. Only at the HUMFOLP23 locus is there an

apparent transversion (A > C); in this case, the mutational event could also be a deletion. The predominance of transitions is also seen at AAAG repeats (Urquhart et al. 1993; Adams et al. 1993) where the usual nonconsensus repeat is AAGG.

#### Non-consensus alleles

Of the 12 loci studied, 6 showed non-consensus alleles, and 2 of these, HUMFOLP23 and HUMCD4, only differed from other alleles in sequence and could not therefore be distinguished by sizing. Nonconsensus alleles showed a distinct tendency towards the ends of allele size ranges. That at HUMCD4 was the second largest allele, while at HUMTH01 the 9.3 allele was 1 bp smaller than the largest common allele. The HUMFOLP23 non-consensus allele was the same size as the largest alleles found, and those at HUMF13A01 and HUMCYAR04 are the smallest alleles found at the loci. Only the extremely rare HUMVWFA31 15.2 allele falls towards the middle of the allele size range.

It is possible that mutation to non-consensus allele is a mechanism which prevents both extreme expansion to high repeat numbers and extreme contraction to low-number, non-polymorphic, repeats. For the 2 loci where the non-consensus allele is at the low end of the size range, HUMF13A01 and HUMCYAR04, deletions outside the repeat sequence cause the non-consensus allele, while those at the top end of their range are caused by either deletion within a repeat (HUMTH01) or substitution within a repeat (HUMFOLP23 and HUMCD4).

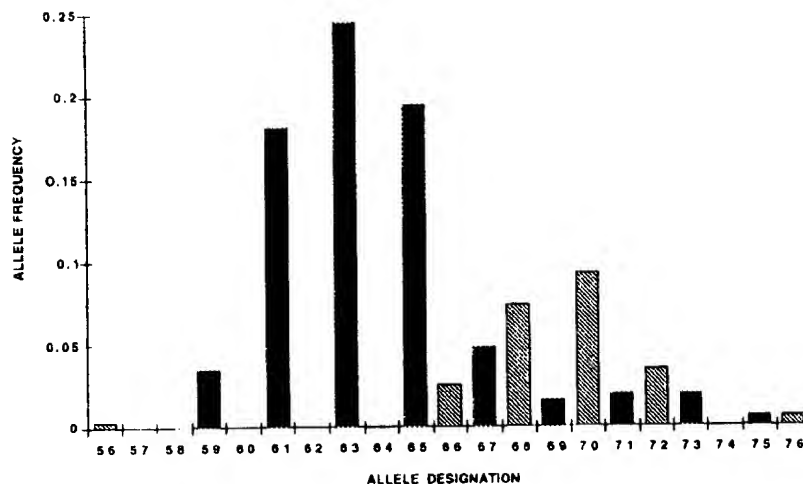
Alleles at the D21S11 locus show a bimodal distribution, with odd-numbered and even-numbered alleles showing distributions over different size ranges (Fig. 4). Presumably this is due to the effect of the TA dinucleotide on increase or decrease of repeat number over time.

Alleles at other polymorphic STR loci require investigation to determine the extent to which sequence effect evolution at these loci.

#### Implications for forensic use

We have surveyed 12 STR loci to investigate candidate loci for an STR system for forensic identification. The major considerations for selection of loci were discriminating power, absence of linkage, agreement with Hardy-Weinberg equilibrium, low levels of 'shadow bands' (Hauge and Litt 1993), compatibility with other loci (for a multiplex STR system) and accurate sizing of alleles. Where alleles differ by 2 bp or more, sizing using the Pst I digest of bacteriophage lambda as a marker consistently distinguished alleles, but alleles differing by 1 bp required sizing by an allelic ladder. Hence, the non-consensus alleles at HUMF13A01 and HUMVWFA31 were sized and designated accurately. However, at the HUMTH01 locus, the 9.3 and 10 alleles were treated as a single pooled allele group, since it was not possible to consistently distinguish

**Fig. 4** Allele distribution at the D21S11 locus. Odd-numbered alleles are solid, even-numbered are hatched, showing two overlapping normal distributions. Data are pooled from 157 British individuals (i.e. 314 chromosomes) or various racial origin



**Table 3** Allele size ranges and non-consensus alleles at the 12 STR loci. Sizes are as determined by sequencing the alleles

Locus	Repeat	Consensus alleles		Non-consensus allele
		Smallest allele	Largest allele	
HUMFES/IPS	ATTT	8 (211 bp)	14 (235 bp)	—
HUMPLA2A1	ATT	8 (110 bp)	17 (137 bp)	—
HUMFABP	ATT	8 (213 bp)	15 (234 bp)	—
HUMTH01	TCAT	5 (154 bp)	11 (178 bp)	9.3 (173 bp)
HUMF13A01	GAAA	4 (183 bp)	17 (235 bp)	3.2 (181 bp)
HUMCYAR04	TTTA	7 (169 bp)	13 (193 bp)	6.1 (166 bp)
HUMFOLP23	AAAM	6 (158 bp)	10 (174 bp)	10 (174 bp)
HUMCD4	YTTTC	3 (96 bp)	13 (146 bp)	12 (141 bp)
HUMGABRB15	KATM	9 (124 bp)	15 (148 bp)	—
HUMTFIIDA	CAR	27 (168 bp)	40 (207 bp)	—
HUMVWFA31	RTCT	12 (130 bp)	21 (166 bp)	15.2 (144 bp)
D21S11	TV <sup>a</sup>	56 (209 bp)	76 (249 bp)	—

<sup>a</sup> Excluding TCA trinucleotide (see text)

between them. This led to a slight loss of informativity at this locus.

The sequence data presented here are of less relevance to non-fluorescent STR analysis in which allele designation is by comparison with an allelic ladder. However, the allele designations suggested here are relevant whichever method of analysis is used. Small (1 or 2 bp) differences in allele sizes can cause problems using non-fluorescent detection methods, particularly where there is an appreciable difference in motility between denatured DNA strands.

In the future, an ideal multiplex STR system would consist of loci in which alleles differ by a minimum of 2 bp. The presence of non-consensus alleles does not rule out loci for inclusion as forensic identification markers, but size differences between alleles of 1 bp require very precise sizing. With the use of allelic ladders, more discriminating hypervariable loci such as HUMACTBP2 (Urquhart et al. 1993) and D11S554 (Adams et al. 1993) could be used. D21S11, though complex in sequence, can be sized in our system and, as a highly discriminating locus, would be a useful component of a multiplex STR system.

From the loci investigated in this study, we have developed a quadruplex STR system including the loci HUM-

FES/FPS, HUMTH01, HUMF13A01, and HUMVWFA31 (Kimpton et al. 1993, 1994), and further, more discriminating, systems are under investigation.

## References

- Adams M, Urquhart A, Kimpton C, Gill P (1993) The human D11S554 locus: four distinct families of repeat pattern alleles at one locus. *Hum Mol Genet* 2: 1373–1376
- Ansorge W, Caskey CT, Erfle H, Zimmermann J, Schwager C, Stegemann J, Civitello A, et al. (1990) Automated DNA sequencing of the human HPRT locus. *Genomics* 6: 593–608
- Dean M, Lucas-Derse S, Bolos A, O'Brien SJ, Kirkness EF, Fraser CM, Goldmann D (1991) Genetic mapping of the b1 GABA receptor gene to human chromosome 4, using a tetranucleotide repeat polymorphism. *Am J Hum Genet* 49: 621–626
- DNA Commission of the International Society for Forensic Haemogenetics (1992) DNA recommendations – 1992 report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. *Int J Leg Med* 105: 63–64

- Edwards MC, Clemens PR, Tristan M, Pizzuti A, Gibbs RA (1991) Pentanucleotide repeat length polymorphism at the human CD4 locus. *Nucleic Acids Res* 19:4791
- Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241-253
- Erickson RP, Ross CE, Gorski JL, Stalvey JR, Drumm MM (1988) Bkm sequences from the human X chromosome contain large clusters of GATA/GACA repeats. *Ann Hum Genet* 52:167-176
- Fregeau CJ, Fournier RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *Bio Techniques* 15:100-109
- Gill P, Sullivan KM, Werrett DJ (1990) The analysis of hypervariable DNA profiles: problems associated with the objective determination of the probability of a match. *Hum Genet* 85:75-79
- Goltsov AA, Eisensmith RC, Naughton EF, Jin L, Chakraborty R, Woo SLC (1993) A single polymorphic STR system in the human phenylalanine hydroxylase gene permits rapid prenatal diagnosis and carrier screening for phenylketonuria. *Hum Mol Genet* 2:577-581
- Hampe A, Shamoon BM, Gobet M, Sherr C, Galibert F (1989) Nucleotide sequence and structural organisation of the human FMS proto-oncogene. *Oncogene Res* 4:9-17
- Hauge XY, Litt M (1993) A study of the origin of 'shadow' bands seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum Mol Genet* 2:411-415
- Hudson TJ, Engelstein M, Lee MK, Ho EC, Rubinfeld MJ, Adams CP, Housman DE, et al. (1992) Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* 113:622-629
- Jackson JA, Fink GR (1981) Gene conversion between duplicated genetic elements in yeast. *Nature* 292:306-311
- Jaenichen HR, Pech M, Lindenmaier W, Wildgruber N, Zachau HG (1984) Composite human V kappa genes and a model of their evolution. *Nucleic Acids Res* 12:5249-5263
- Jeffreys AJ, Macleod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354:204-209
- Khan AS, Wilcox AS, Polymeropoulos MH, Hopkins JA, Stevens TJ, Robinson M, Orpana AK, et al. (1992) Single pass sequencing and physical and genetic mapping of human cDNAs. *Nature Genet* 2:180-185
- Kimpton CP, Walton A, Gill P (1992) A further tetranucleotide repeat polymorphism in the vWF gene. *Hum Mol Genet* 1:287
- Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Applic* 3:13-22
- Kimpton CP, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, Gill P (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *Int J Leg Med* (in press)
- Mahtani MM, Willard HF (1993) A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci. *Hum Mol Genet* 2:431-437
- Mancuso DJ, Tuley EA, Westfield LA, Worrall NK, Shelton-Inloes BB, Sorace JM, Alevy YG, et al. (1989) Structure of the gene for human von Willebrand factor. *J Biol Chem* 264:19514-19527
- Murphy S, Altruda F, Ullu E, Tripodi M, Silengo L, Melli M (1984) DNA sequence complementary to human 7 SK RNA show structural similarities to the short mobile elements of the mammalian genome. *J Mol Biol* 177:575-590
- Nomenclature Committee of the International Union of Biochemistry (1985) Nomenclature for incompletely specified bases in nucleic acid sequences. *Recommendations 1984. Eur J Biochem* 150:1-5
- Phromchotikul T, Browne D, Litt M (1992) Microsatellite polymorphisms at the D11S554 and D11S569 loci. *Hum Mol Genet* 1:214
- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1990a) Trinucleotide repeat polymorphism at the human intestinal fatty acid binding protein gene (FABP2). *Nucleic Acids Res* 18:7198
- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1990b) Trinucleotide repeat polymorphism at the human pancreatic phospholipase A-2 gene (PLA2). *Nucleic Acids Res* 18:7468
- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1991a) Tetranucleotide repeat polymorphism at the human aromatase cytochrome P-450 gene (CYP19). *Nucleic Acids Res* 19:195
- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1991b) Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). *Nucleic Acids Res* 19:4018
- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1991c) Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). *Nucleic Acids Res* 19:4036
- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1991d) Tetranucleotide repeat polymorphism at the human dihydrofolate reductase psi-2 pseudogene (DHFRP2). *Nucleic Acids Res* 19:4792
- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1991e) Trinucleotide repeat polymorphism at the human transcription factor IID gene. *Nucleic Acids Res* 19:4037
- Polymeropoulos MH, Xiao H, Rath DS, Merrill CR (1991f) Tetranucleotide repeat polymorphism at the human tyrosine hydrolase gene (TH). *Nucleic Acids Res* 19:3753
- Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTHO1[AATG]<sub>n</sub> and reassignment of alleles in population analysis by using a locus-specific allelic ladder. *Am J Hum Genet* 53:953-958
- Riley DE, Goldman MA, Gantler SM (1991) Nucleotide sequence of the 3' nuclease-sensitive region of the human phosphoglycerate kinase (PGD1) gene. *Genomics* 11:212-214
- Roewer L, Arnemann J, Spurr NK, Grzeschik K-H, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum Genet* 89:389-394
- Sharma V, Litt M (1992) Tetranucleotide repeat polymorphism at the D21S11 locus. *Hum Mol Genet* 1:67
- Slightom JL, Koop BF, Xu P, Goodman M (1988) Rhesus fetal globin genes. *J Biol Chem* 263:12427-12438
- Urquhart AJ (1991) Human tyrosinase-like protein (TYRL) carboxy terminus: closer homology with the mouse protein than previously reported. *Nucleic Acids Res* 19:5803
- Urquhart A, Kimpton CP, Gill P (1993) Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. *Hum Genet* 92:637-638
- Warne D, Watkins C, Bodfish P, Nyberg K, Spurr N (1991) Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene 2 (ACTBP2) detected using the polymerase chain reaction. *Nucleic Acids Res* 19:6980
- Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388-396
- Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, et al. (1992) A second-generation linkage map of the human genome. *Nature* 359:794-801
- Wiegand P, Budowle B, Rand S, Brinkmann B (1993) Forensic validation of the STR systems SE33 and TC11. *Int J Leg Med* 105:315-320
- Yu L, Peng C, Starnes SM, Liou RS, Chang T-W (1990) Two isoforms of human membrane-bound alpha Ig resulting from alternative mRNA splicing in the membrane segment. *J Immunol* 145:3932-3936

**STIC-ILL**

*mic*  
*PR355, J65*

**From:** Goldberg, Jeanine  
**Sent:** Friday, December 01, 2000 2:57 PM  
**To:** STIC-ILL  
**Subject:** please pull HCV

1. Proc. Natl. Acad. Sci. U.S.A. 91 (21), 10134-10138 (1994)
2. J. Virol. 66, 3225-3229 (1992)
3. Hepatology 19 (3), 551-553 (1994)
4. J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)

THANK YOU

Jeanine Enewold Goldberg  
1655  
CM1--12D11  
306-5817

## Hepatitis C Virus (HCV) Circulates as a Population of Different but Closely Related Genomes: Quasispecies Nature of HCV Genome Distribution

MARIA MARTELL,<sup>1</sup> JUAN I. ESTEBAN,<sup>1\*</sup> JOSEP QUER,<sup>1</sup> JOAN GENESCÀ,<sup>1</sup> AMY WEINER,<sup>2</sup>  
RAFAEL ESTEBAN,<sup>1</sup> JAUME GUARDIA,<sup>1</sup> AND JORDI GÓMEZ<sup>1</sup>

*Liver Unit, Department of Medicine, Hospital General Universitari Vall d'Hebron,  
08035 Barcelona, Spain,<sup>1</sup> and Chiron Corporation,  
Emeryville, California 94608-2916<sup>2</sup>*

Received 22 November 1991/Accepted 20 January 1992

Sequencing of multiple recombinant clones generated from polymerase chain reaction-amplified products demonstrated that the degree of heterogeneity of two well-conserved regions of the hepatitis C virus (HCV) genome within individual plasma samples from a single patient was consistent with a quasispecies structure of HCV genomic RNA. About half of circulating RNA molecules were identical, while the remaining consisted of a spectrum of mutants differing from each other in one to four nucleotides. Mutant sequence diversity ranged from silent mutations to appearance of in-frame stop codons and included both conservative and nonconservative amino acid substitutions. From the relative proportion of essentially defective sequences, we estimated that most circulating particles should contain defective genomes. These observations might have important implications in the physiopathology of HCV infection and underline the need for a population-based approach when one is analyzing HCV genomes.

Hepatitis C virus (HCV), a 10-kb positive-stranded RNA virus, has recently been shown to be the major causative agent of parenterally transmitted non-A, non-B hepatitis (1, 5, 11, 21). Despite little overall primary sequence identity, the genetic organization of HCV has been shown to be similar to those of flaviviruses and pestiviruses (6, 26, 33). Significant genetic heterogeneity has been reported among isolates from different geographic areas (17) and within single isolates from the same individual (17, 20, 33-35). Comparative sequence analysis of the different HCV isolates has shown, however, that the degree of variability is unevenly distributed throughout the HCV genome, with some very well conserved regions (6, 13, 17, 33-35) and some highly variable genes (17, 20, 33-35). In addition, the rate of fixation of mutations of the HCV genome has been estimated to be similar to that of other RNA viruses (approximately  $10^{-3}$  to  $10^{-4}$  base substitutions per genome site per year), and evidence suggesting that the HCV genome may rapidly evolve in vivo, with different rates of evolution for different regions of the viral genome, has been provided (27).

Since the pioneering studies by Batschelet et al. (3) providing evidence that RNA virus heterogeneity is a consequence of high error rates in RNA replication, data have accumulated which suggest that most RNA viruses consist of a heterogeneous mixture of circulating related genomes containing a master (most frequently represented) sequence and a large spectrum of mutants, a genomic distribution referred to as quasispecies (10, 32). This quasispecies model of mixed RNA virus populations implies a significant adaptation advantage because the simultaneous presence of multiple variant genomes (and the high rate of generation of new variants) allows for the rapid selection of the mutant(s) with better fitness for any new environmental condition. On the other hand, a quasispecies will remain in stable equilibrium

with little evolution of its consensus or master sequence as long as conditions are unchanged (7, 15, 18, 22).

Many important biological implications predicted by the quasispecies model (8) have been found in several virus systems displaying such genomic structure, including vaccination failure through selection of neutralizing antibody escape mutants (19), establishment of persistent infection by selection of neutralizing antibody or cytotoxic T-lymphocyte escape mutants (23, 28, 30) or by generation of defective interfering particles (14), resistance to antiviral agents (29), and changes in cell tropism or virulence (9, 12).

In an attempt to define the degree of heterogeneity among circulating HCV RNA molecules within individual isolates, we analyzed a series of four sequential HCV isolates from an individual with transfusion-associated HCV and human immunodeficiency virus coinfection. Polymerase chain reaction (PCR)-amplified products of cDNA corresponding to fragments of the 5' untranslated region (5'UTR) and non-structural region 3 (NS3) of the HCV genome were cloned into a bacterial vector, and 6 to 20 recombinant clones from each sample were sequenced. A quasispecies distribution of sequences was observed in both genomic regions and in all sequential samples.

HCV RNA extracted from 75  $\mu$ l of plasma by the acid guanidinium thiocyanate-phenol-chloroform method (4) was reverse transcribed into cDNA and PCR amplified in a single tube reaction (RT-PCR; kit N 808-007, Perkin Elmer) for 35 cycles (5 cycles of 94°C for 2 min, 50°C for 2 min, and 72°C for 3 min; 30 cycles of 94°C for 1.5 min, 60°C for 2 min, and 72°C for 3 min), using specific oligonucleotide primers of the 5'UTR (13) and NS3/NS4 regions (16). As represented in Fig. 1, the amplified products were 237 and 584 bp long, and their specificity was confirmed by Southern hybridization with <sup>32</sup>P probes. 5'UTR and NS3/NS4 products were cleaved with *NorI* and *SacI-NcoI*, respectively, yielding restriction products of 177 and 240 bp, respectively, which were subsequently cloned into pSL1190 vector (Pharmacia-

\* Corresponding author.

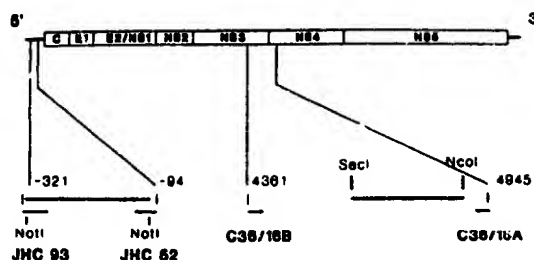


FIG. 1. Organization of the HCV genome. Putative structural genome regions: core (C), envelope (E1 and E2/NS1), nonstructural protein coding regions (NS1 to NS5). Amplified 5'UTR and NS3/NS4 products are depicted below, with their respective ends indicated by their corresponding nucleotide positions from the HCV-1 prototype (6, 13). Horizontal solid lines represent restriction fragments generated from each product for subsequent cloning and sequencing. Oligonucleotide primers JHC 52 (5'-AGTCTTGGCGC CGCACGCGCCAAATC), JHC 93 (5'-TTCGCGGCCGCACTC CACCATGAATCACTCCCC), C36/16A (5'-GCATGTCATGATGT AT), and C36/16B (5'-GCAATACGTGTGTCAC) used for cDNA synthesis and PCR amplification were synthesized according to the HCV-1 prototype sequence (13, 16).

1.KB). Positive recombinant clones were subcloned into M13mp19. Six to twenty independent recombinant subclones were sequenced by the dideoxy-mediated chain termination method (31). The Maxam-Gilbert chemical degradation method (24) was used to resolve sequencing ambiguities.

When the sequences of 20 independent NS3 clones from the sample obtained 2 months posttransfusion were compared (Table 1 and Fig. 2a), 9 sequences (45%) were identical (master sequence) and 11 (55%) differed from each other and from the master sequence in one to four nucleotides. Similarly, of the 20 5'UTR clones from the same sample, 12 sequences (60%) were identical and 8 (40%) contained single, nonrepetitive base substitutions (Table 1 and Fig. 2b). The predicted amino acid sequence encoded by the master NS3 sequence was identical to that of the corresponding HCV-1 prototype (16). However, of the 25 base substitutions observed (Fig. 2a) in the 11 mutant sequences, only 5 (20%), occurring in the third codon position, were silent mutations. Of the 20 nucleotide changes observed in the first and second codon positions, 6 (24%) led to conservative amino acid changes, 11 (44%) introduced drastic amino acid changes (five Ile → Thr, three Leu → Arg, one Asp → Val, one Met → Thr, and one Tyr → His), and 3 (12%) led to the appearance of in-frame stop codons in 2 of the 20 sequences. Although the functional implications of single amino acid substitutions (whether conservative or nonconservative) are unknown, it is obvious that mutants containing in-frame stop codons are essentially defective. It must be borne in mind

that the sequenced fragment of the NS3 region encodes only 80 of a total of 3,011 residues (i.e., 2.8% of the coding capacity of the HCV genome). Assuming that the NS3 region is representative of the HCV open reading frame, the finding that 10% of these sequences include stop codons indicates that most circulating particle populations must contain defective genomes.

To ensure that the observed heterogeneity was not due to nucleotide misincorporations introduced by the reverse transcriptase or the *Taq* DNA polymerase during the amplification reaction, two control experiments were carried out. First, one of the NS3 recombinant clones of known sequence was in vitro transcribed with SP6 RNA polymerase (Boehringer Mannheim). The RNA transcript was subjected to the original RT-PCR amplification procedure under identical conditions, and the amplified product was subcloned. Sequence analysis of five independent clones showed absolute identity with the parental clone from which the RNA transcript had been obtained. In a second experiment, the cDNA insert of one of the sequenced 5'UTR recombinant clones was subjected to 35 additional cycles of PCR amplification and subcloned. Within 23 such clones (5,451 bases sequenced), only one nucleotide change (A → G) was detected in one of the clones. Therefore, it seems that the observed heterogeneity was not an artifact, although some of the base substitutions might have been introduced during the amplification procedure.

With such a low level of misincorporation noise, the finding that two distinct and well-conserved regions of the viral genome (17) were composed of a mixture of cocirculating related genomes distributed as a predominant or master sequence and a spectrum of mutants (representing altogether about half of the genomic population) provides unequivocal evidence of the quasispecies structure of the HCV genome. Additional sequencing of 6 to 10 independent 5'UTR recombinant clones, generated from subsequent samples obtained later within the acute phase (at 3 months posttransfusion) and during the chronic phase (at 6 and 29 months posttransfusion), showed similar distribution and complexity of the quasispecies for that region again with randomly distributed single nonrepetitive base substitutions in each of the mutants around a conserved master sequence (Table 1). This finding demonstrates that the observed quasispecies distribution was not a phenomenon restricted to the acute phase of the infection.

Demonstration of the quasispecies structure of the HCV genome implies the need for a population-based approach when one is analyzing HCV genomes. Accordingly, the average or consensus sequence of each nucleic acid region should always be carefully defined when one is trying to determine the complete genomic sequence of a given isolate. Otherwise, reconstruction of the HCV genome by randomly sequencing overlapping clones of PCR products obtained from plasma pools of different subjects (or even from a single patient) would generate an artifactual genomic mosaicism.

Given our limited knowledge of the biological functions of most HCV genome regions and of their encoded products, besides the potential problems for vaccine development, it is not possible at present to anticipate which of the several relevant implications predicted by the quasispecies model will be applicable to HCV. One may speculate, however, on two possible mechanisms which might play a role in the two most striking features of HCV infection: its high tendency toward viral persistence and the characteristic smoldering and fluctuating course of chronic hepatitis C. On the one hand, the high degree of heterogeneity of the envelope-

TABLE 1. Sequence complexity of HCV quasispecies

Sample	Posttransfusion mo	Genomic region cDNA	No. of clones sequenced	No. of identical nucleic acid sequences	No. of different nucleic acid sequences
VL-1	2	NS3	20	9	11
		5'UTR	20	12	8
VL-2	3	5'UTR	10	7	3
VL-3	6	5'UTR	10	7	3
VL-4	29	5'UTR	6	4	2

**a** NS3/NS4 REGIONClone  
number

4814

L R A Y N H T P G L P V C O D L E F U E S V F T G L T I D A H  
 5 GCTACGAGCGTACATGAACACTCCGGGACTTCCCGTGTCCCAAGACCATCTTGAATTTTGGGAGGCGCTTTACGGGTCTCAACCATATAGAGGCGGAC  
 T  
 11 -----C-----  
 T  
 10 -----C-----  
 4 -----  
 7 (H) -----  
 C T  
 15 -----C-----  
 17 -----T----- (H) 1  
 3 (H) L L -----  
 C T C  
 19 -----  
 8 (H) " " -----  
 G T A  
 2 (H) " " -----  
 G T

F L S D T E Q S G E N L P T L V A V G A T V C A R A A A P P S U D  
 5 TTCTATGCCAACAAGCAGACCGGGGAACCTTCTTACCTGGTAGCGTATCAAGCCACCGTGTGCGCTAGAGCCCAAGCTCTCTCCCGCTGGG  
 11 -----  
 10 -----  
 4 -----  
 7 -----  
 15 -----C-----  
 17 -----  
 3 -----  
 19 -----  
 8 -----  
 2 -----

Q H N E C L I R L E P T L 4854  
 5 ACCAGATGGGAATGCTTGATCGCTCTCAAGCCACCTC 45%  
 (T)  
 11 -----C----- 10%  
 10 -----E----- 5%  
 4 -----A----- 5%  
 7 X ----- 5%  
 G  
 15 -----C----- 5%  
 17 ----- 5%  
 3 (T) ----- 5%  
 (T) (T) ----- 5%  
 19 -----C----- 5%  
 8 B ----- 5%  
 G  
 2 Y -----C----- 5%  
 (T)

**b** NSV 5'UTR REGIONClone  
number

4814

1 TGTGAGGAAGTACTGTCTTCAAGCAAAAGCTGTACCCATGCGCTTATATGATGTCTGACAGCTCCAGGACCCCTCCCGGAGAGCCATAGT  
 34 -----  
 38 -----G-----  
 42 -----G-----  
 54 -----A-----  
 8 -----T-----  
 74 -----G-----  
 55 -----T-----T-----  
 4 -----A-----

1 GTCTGGGAAGCGTGAATGACAGCAAAAGCTGTACCCATGCGCTTATATGATGTCTGACAGCTCCAGGACCCCTCCCGGAGAGCCATAGT 60%  
 34 -----G----- 5%  
 38 ----- 5%  
 42 ----- 5%  
 54 ----- 5%  
 8 ----- 5%  
 74 ----- 5%  
 55 ----- 5%  
 4 ----- 5%

FIG. 2. Nucleotide and amino acid sequences of 20 independent NS3/NS4 (a) and 5'UTR (b) clones from sample VL-1 (see Table 1). Relative frequencies are shown on the right. Mutant sequences have been aligned to the master (most frequently represented) sequence for each region. The one-letter code has been used for protein sequences. Dots indicate sequence identity. Only nucleotides which differ from the master sequence are shown. Asterisks indicate in-frame stop codons. Conservative amino acid substitutions are underlined; those in brackets correspond to nonconservative changes.

encoding regions and its potential for rapid evolution (as anticipated by the quasispecies model) might lead to rapid antigenic changes in structural proteins likely to contain neutralizing antibody and/or cytotoxic T-cell-specific epitopes and hence provide the virus the opportunity to avoid the previous response of the host immune system. On the other hand, if our assumption that most circulating viral particles contain defective genomes is true, as it has been postulated for other viruses (25), this would provide an alternative or supplementary mechanism for HCV persistence. Indeed, there is a wealth of data indicating how defective RNA genomes may restrict viral replication in vivo, modulate the clinical expression of the disease, and lead to the establishment of persistent viral infection (2, 14, 18). Whether, or the extent to which, these mechanisms play a role in the natural history of HCV infection remains to be determined. The widespread nature of chronic HCV infection and its long-term clinical consequences, however, warrant further studies on any potential mechanism of HCV persistence.

We are indebted to E. Domingo (Centro de Biología Molecular, CSIC, Universidad Autónoma de Madrid) for continuous encouragement and support and useful discussion of the manuscript. We also thank P. Puigdomenech (Centro de Investigación y Desarrollo, CSIC, Barcelona, Spain) for useful scientific and technical support. We thank J. H. Han (Chiron Corporation, Emeryville, Calif.) for kindly providing primers JHC 93 and JHC 52.

This work was supported in part by grants SAL 90/0449 from the Comisión Interministerial de Ciencia y Tecnología and 91/0298 from the Fondo de Investigaciones Sanitarias.

#### REFERENCES

- Alter, H. J., R. H. Purcell, J. W. K. Shih, J. C. Melpolder, M. Houghton, Q. L. Choo, and G. Kuo. 1989. Detection of antibody to hepatitis C virus in prospectively followed transfusion-recipients with acute and chronic non-A non-B hepatitis. *N. Engl. J. Med.* 321:1494-1500.
- Barret, A. D. T., and N. J. Dimmock. 1986. Defective interfering viruses and infections of animals. *Curr. Top. Microbiol. Immunol.* 128:55-84.
- Batschelet, E., E. Domingo, and C. Weissmann. 1976. The proportion of revertant and mutant phage in a growing population as a function of mutation and growth rate. *Gene* 1:27-32.
- Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156-159.
- Choo, Q. L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244:359-362.
- Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, D. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Bar, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton. 1991. Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* 88:2451-2455.
- Domingo, E. 1989. RNA virus evolution and the control of viral disease. *Prog. Drug Res.* 33:93-133.
- Domingo, E., E. Martínez-Salas, F. Sobrino, J. C. de la Torre, A. Portela, J. Ortín, C. López-Galíndez, P. Pérez-Breña, N. Villanueva, R. Nájera, S. VandePol, N. DePolo, and J. Holland. 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. *Gene* 40:1-8.
- Domingo, E., M. G. Mateu, M. A. Martínez, J. Dopazo, A. Moya, and F. Sobrino. 1990. Genetic variability and antigenic diversity of foot-and-mouth disease virus. p. 233-266. *In* E. Kurstak, R. G. Marusyk, F. A. Murphy, and M. H. V. Van Regenmortel (ed.), *Applied virology research*, vol. 2. Plenum Publishing Corp., New York.
- Domingo, E., D. Sabo, T. Taniguchi, and C. Weissmann. 1978. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13:735-744.
- Esteban, J. I., R. Esteban, L. Viladomiu, J. C. López-Talavera, A. González, J. M. Hernández, M. Roget, J. Genesca, M. Buti, G. Kuo, Q. L. Choo, M. Houghton, and J. Guardia. 1989. Hepatitis C virus antibodies among risk groups in Spain. *Lancet* ii:294-297.
- Evans, D. M. A., G. Dunn, P. D. Minor, G. C. Schild, A. J. Cann, G. Stanway, J. W. Almond, K. Currey, and J. V. Maizel, Jr. 1985. Increased neurovirulence associated with single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome. *Nature (London)* 314:548-550.
- Han, J. H., V. Shyamala, K. H. Richman, M. J. Brauer, B. Irvine, M. S. Urdea, P. Tekamp-Olsen, G. Kuo, Q. L. Choo, and M. Houghton. 1991. Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5' untranslated region and poly(A) tails at the 3' end. *Proc. Natl. Acad. Sci. USA* 88:1711-1715.
- Holland, J. J. 1991. Defective viral genomes, p. 151-165. *In* B. N. Fields and D. N. Knipe (ed.), *Fundamental virology*. Raven Press, New York.
- Holland, J. J., J. C. de la Torre, D. K. Clark, and E. Duarte. 1991. Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *J. Virol.* 65:2960-2967.
- Houghton, M., Q. L. Choo, and G. Kuo. March 1989. European patent application 88 310 922.5; publication 318.216.
- Houghton, M., A. J. Weiner, J. H. Han, G. Kuo, and Q. L. Choo. 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. *Hepatology* 14:381-388.
- Huang, A. S. 1988. Modulation of viral disease processes by defective interfering particles, p. 195-208. *In* E. Domingo, J. J. Holland, and P. Ahlquist (ed.), *RNA genetics*, vol. 3. CRC Press, Boca Raton, Fla.
- Kew, O. M., and B. K. Nottay. 1984. Evolution of the oral polio vaccine strains in humans occurs by both mutations and intramolecular recombination, p. 357-367. *In* R. M. Chanock and R. A. Lerner (ed.), *Modern approaches to vaccines*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Kremsdorff, D., C. Porchon, J. P. Kim, G. R. Reyes, and C. Bréchet. 1991. Partial nucleotide sequence analysis of a French hepatitis C virus: implications for HCV genetic variability in the E2/NS1 protein. *J. Gen. Virol.* 72:2557-2561.
- Kuo, G., Q. L. Choo, H. J. Alter, G. L. Gitnick, A. G. Redeker, R. H. Purcell, T. Miyamura, J. L. Dienstag, M. J. Alter, C. E. Stevens, G. E. Tegtmeier, F. Bonino, M. Colombo, W. S. Lee, C. Kuo, K. Berger, J. R. Schuster, L. R. Overby, D. W. Bradley, and M. Houghton. 1989. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* 244:362-364.
- Martínez, M. A., C. Carrillo, F. González-Candenas, A. Moya, E. Domingo, and F. Sobrino. 1991. Fitness alteration of foot-and-mouth disease virus mutants: measurement of adaptability of viral quasispecies. *J. Virol.* 65:3954-3957.
- Mateu, M. G., M. A. Martínez, E. Rocha, D. Andreu, J. Parejo, E. Giralt, F. Sobrino, and E. Domingo. 1989. Implications of a quasispecies genome structure: effect of frequent naturally occurring amino acid substitutions on the antigenicity of foot-and-mouth disease virus. *Proc. Natl. Acad. Sci. USA* 86:5863-5867.
- Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74:560-564.
- Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt-Manson, B. Asjö, and S. Wain-Hobson. 1989. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell* 58:901-910.
- Miller, R. H., and R. H. Purcell. 1990. Hepatitis C virus shares amino acid sequence similarity with pestivirus and flavivirus as well as members of two plant virus supergroups. *Proc. Natl. Acad. Sci. USA* 87:2057-2061.
- Ogata, N., H. J. Alter, R. H. Miller, and R. H. Purcell. 1991. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc. Natl. Acad. Sci. USA* 88:3392-3396.

28. Parry, N., G. Fox, D. Rowlands, F. Brown, E. Fry, R. Acharya, D. Logan, and D. Stuart. 1990. Structural and serological evidence for a novel mechanism of antigenic variation in foot-and-mouth disease virus. *Nature (London)* 347:569-572.
29. Remington, K. M., B. Chesebro, K. Wehrly, N. C. Pedersen, and T. W. North. 1991. Mutants of feline immunodeficiency virus resistant to 3'-azido-2'-deoxythymidine. *J. Virol.* 65:308-312.
30. Salvato, M., P. Borrow, E. Shimomaye, and M. B. A. Oldstone. 1991. Molecular basis of viral persistence: a single amino acid change in the glycoprotein of lymphocytic choriomeningitis virus is associated with suppression of the antiviral cytotoxic T-lymphocyte response and establishment of persistence. *J. Virol.* 65:1863-1869.
31. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
32. Steinhauer, D. A., and J. J. Holland. 1987. Rapid evolution of RNA viruses. *Annu. Rev. Microbiol.* 41:409-433.
33. Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida, and H. Okayama. 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. *J. Virol.* 65:1105-1113.
34. Takeuchi, K., Y. Kubo, S. Boonmar, Y. Watanabe, T. Katayama, Q. L. Choo, G. Kuo, M. Houghton, I. Saito, and T. Miyamura. 1990. The putative nucleocapsid and envelope protein genes of hepatitis C virus determined by comparison of the nucleotide sequences of two isolates derived from an experimentally infected chimpanzee and healthy human carriers. *J. Gen. Virol.* 71:3027-3033.
35. Weiner, A. J., M. J. Brauer, J. Rosenblatt, K. H. Richman, J. Tung, K. Crawford, F. Bonino, G. Saracco, Q. L. Choo, M. Houghton, and J. H. Han. 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the Flavivirus envelope and NS1 proteins and Pestivirus envelope glycoproteins. *Virology* 180:842-848.

**STIC-ILL**

AB43.7.1463  
ADONIS

**From:** Goldberg, Jeanine  
**Sent:** Friday, December 01, 2000 3:12 PM  
**To:** STIC-ILL  
**Subject:** please pull

1. JOURNAL OF VIROLOGICAL METHODS, (1991 Dec) 35 (3) 297-304.  
Journal code: HQR. ISSN: 0166-0934.
2. MOLECULAR AND CELLULAR PROBES, (1993 Oct) 7 (5) 405-10.  
Journal code: NG9. ISSN: 0890-8508.

THANK YOU

Jeanine Enewold Goldberg  
1655  
CM1--12D11  
306-5817

## HIV-1 and HCV co-infected patients: detection of active viral expression using a nested polymerase chain reaction

L. Richard,<sup>1</sup> J.-L. Pellegrin,<sup>2</sup> P. Barbeau,<sup>2</sup> G. Brossard,<sup>2</sup> B. Leng<sup>2</sup> and H. J. A. Fleury<sup>1\*</sup>

<sup>1</sup>Laboratoire de Virologie, Université de Bordeaux II, 33076 Bordeaux, France, and <sup>2</sup>Service des Maladies Infectieuses, Hôpital du Haut-Lévêque, 33600 Pessac, France

(Received 12 May 1993, Accepted 22 June 1993)

The aim of the present study was to determine, in a population of 70 HIV-1 infected patients with antibodies to HCV, the percentages of individuals with an active replication of HIV-1, HCV or both. During a one year follow-up of these patients at different stages of disease, blood samples were regularly collected for determination of transaminases,  $\beta$ 2 microglobulin and CD4<sup>+</sup> lymphocytes. Total RNAs were extracted from the sera, retrotranscribed with MoMuLV reverse transcriptase and nested PCR assays were carried out separately with sets of primers homologous to the 5' non-translated region of HCV and in HIV-1 gag. The amplified products were subjected to electrophoresis and observed under u.v. illumination after staining with ethidium bromide. For some samples, the identity of the amplified products was confirmed by Southern blotting by hybridization with enzyme-labelled probes.

A total of 57% of the patients were found to produce HIV-1 RNA and 62% HCV RNA, while 34% produced both. HIV-1 RNA production was correlated with  $\beta$ 2 microglobulin and CD4<sup>+</sup> levels; active replication of HCV was correlated with hepatitis but not with CD4<sup>+</sup> levels nor with HIV-1 RNA synthesis.

**KEYWORDS:** HIV-1, HCV, RNA, polymerase chain reaction.

### INTRODUCTION

Some years ago, HIV-1 infected patients were considered to have active viral transcription and viral synthesis while entering stage IV of the CDC classification; these data were mainly based on the detection of HIV-1 antigenemia (HIV-1 Ag). More recently, virus titration in the plasma,<sup>1–6</sup> detection of viral RNA in peripheral blood mononuclear cells (PBMC) and/or plasma by polymerase chain reaction (PCR) after a reverse transcription (RT) step<sup>7–14</sup> or *in situ* hybridization (ISH) with RNA probes<sup>15</sup> has shown that asymptomatic patients (at stage II of the CDC) are likely to transcribe viral and/or messenger HIV-1 RNAs and to produce infectious virus particles. These data can

have therapeutic implications, since it now seems that there is no real state of latency after primary infection.

Some of these patients, mainly intravenous drug users, can be co-infected with hepatitis C virus (HCV). HCV primary infection is frequently followed by a chronic state. While the detection of anti-HCV antibodies is useful for diagnosis, they cannot provide reliable data as to the level of viral synthesis.<sup>16,17</sup> This synthesis can be determined by a RT-PCR for the detection of HCV-RNA.<sup>18–22</sup>

The aim of the present study was to evaluate, in a population of 70 HIV-1 infected patients exhibiting

\* Author to whom correspondence should be addressed.

antibodies to HCV, the percentage of individuals with active synthesis of HIV-1, HCV and both, as detected by RT PCR.

## MATERIALS AND METHODS

One hundred and nineteen HIV-1 infected patients (Western blot confirmed) at different stages of the disease, consecutively hospitalized between 1 October 1991 and 31 December 1991, were screened for the presence of antibodies to HCV (Diagnostics Pasteur EIA and Chiron RIBA confirmation); 70 of them were found to be HIV-1/HCV co-infected (52 IV drug abusers; 18 blood transfused). Whole blood samples were collected from these patients; CD<sup>4</sup> lymphocytes were numbered by flow cytometry. HIV-1 Ag was determined by Diagnostics Pasteur EIA. Serum  $\beta$ 2 microglobulin was titrated using Behring EIA. Transaminase levels were determined and chronic hepatitis was defined as transaminase levels higher than 1.5 times the normal level for more than 6 months. Aliquots of serum were immediately frozen at  $-80^{\circ}\text{C}$  before being used for PCR.

### PCR for detection of HIV-1 and HCV RNAs

#### Preparation of RNA

RNA from 100  $\mu\text{l}$  of serum was extracted using 0.5 ml 4 M guanidium isothiocyanate, 25 mM sodium acetate, pH 7, 0.5% sarcosyl, 7%  $\beta$ -mercaptoethanol, 50  $\mu\text{l}$  2 M sodium acetate, pH 4, 0.5 ml phenol, 100  $\mu\text{l}$  chloroform/isoamylalcohol (49:1). The mixture was vigorously shaken, then allowed to stay for 15 min on ice before being centrifuged at 12,000 g for 30 min at  $4^{\circ}\text{C}$ . This step of extraction was followed by precipitation: 500  $\mu\text{l}$  of the supernatant was added to 0.5 ml isopropanol with 50  $\mu\text{l}$  3 M sodium acetate, pH 7, and 5  $\mu\text{l}$  glycogen, then allowed to precipitate at  $-20^{\circ}\text{C}$  for 2 h before centrifugation at 12,000 g for 30 min. The supernatant was discarded while the pellet was washed with 2 ml of 75% ethanol and centrifuged for 15 min at 12,000 g. The pellet was dried and dissolved in 50  $\mu\text{l}$  RNase-free  $\text{H}_2\text{O}$ , heated at  $65^{\circ}\text{C}$  for 20 min before being frozen at  $-80^{\circ}\text{C}$ .

#### Treatment with RNase-free DNase

A total of 10  $\mu\text{l}$  of RNA was treated during 1 h at room temperature with 30 u of RNase-free DNase; the reaction was stopped by heating for 5 min at  $94^{\circ}\text{C}$ .

### Reverse transcription

Reverse transcription was carried out for 1 h at  $42^{\circ}\text{C}$  with 2  $\mu\text{l}$  10 $\times$  buffer 500 mM KCl, 100 mM Tris-HCl, pH 8.3, 15 mM  $\text{MgCl}_2$ , 0.1% gelatin, 1  $\mu\text{l}$  dNTP (10 mM each), 1  $\mu\text{l}$  virus-specific antisense primer (100 pM) of the first PCR (see below), 1  $\mu\text{l}$  RNasin (20 u), 1  $\mu\text{l}$  50 mM  $\text{MgCl}_2$ , 3  $\mu\text{l}$   $\text{H}_2\text{O}$ , 1  $\mu\text{l}$  MoMuLV (Moloney Murine Leukaemia Virus) reverse transcriptase (100–200 u) and 10  $\mu\text{l}$  RNA. The reaction was stopped by heating at  $95^{\circ}\text{C}$  for 5 min. HIV-1 and HCV PCR were then performed using a nested technique.

### Nested PCR for HIV-1

Two sets of primers were used to amplify a conserved region of HIV-1 gag (Fig. 1). The outer primers were 5' ATAATCCACCTATCCCAGTAGGAG 3' (SK38) (sense) and 5' TGACATGCTGTCATCATTCTTC 3' (antisense); the inner primers were SK 38 (sense) and 5' TTGGTCCTGTCTTATGTCCAGAATGC 3' (SK 39) (antisense).

To 50  $\mu\text{l}$  of the reaction mixture was added the cDNA, 30 pM of the primers, 10 mM of each dNTP, 10 mM Tris-HCl, pH 8.3, 1.5 mM  $\text{MgCl}_2$ , 50 mM KCl and 2.5 u of Taq polymerase (Boehringer). The mixture was then overlaid with 50  $\mu\text{l}$  of mineral oil and the tubes were placed in a DNA thermal cycler (Perkin-Elmer Cetus).

The parameters of both rounds of amplification were identical: 10 min at  $94^{\circ}\text{C}$  followed by 35 cycles of 1 min at  $94^{\circ}\text{C}$ , 1 min at  $51^{\circ}\text{C}$  and 1 min at  $72^{\circ}\text{C}$ . Ten microlitres of the amplified products of both rounds corresponding to 276 and 115 bp were subjected to electrophoresis in a 3% agarose gel (Nusieve, Sigma) and stained with ethidium bromide. The presence of a 115 bp band in the second round was considered a positive result.

### Nested PCR for HCV

All used primers had been selected to amplify the highly conserved 5' non-translated region of the HCV genome.<sup>23,24</sup> The outer primers were: 5' TGGGGGCGACACTCCACCATAGAT 3' (sense) and 5' CGTGCTGGTGACGGGTGTACGAGACCT -3' (antisense) and the parameters of the first round were as follows: 10 min at  $94^{\circ}\text{C}$  then 35 cycles of 1 min at  $94^{\circ}\text{C}$ , 2 min of annealing at  $50^{\circ}\text{C}$  and 3 min at  $72^{\circ}\text{C}$ .

The inner primers were: 5' CCACCATAGATCACTCCCTGT 3' (sense) and 5' CACTCGCAAGCACCCTATCAGGCAGT 3' (antisense). The parameters were 10 min at  $94^{\circ}\text{C}$  followed by 35 cycles of 1 min at  $94^{\circ}\text{C}$ , 2 min at  $46^{\circ}\text{C}$  and 3 min at  $72^{\circ}\text{C}$ .

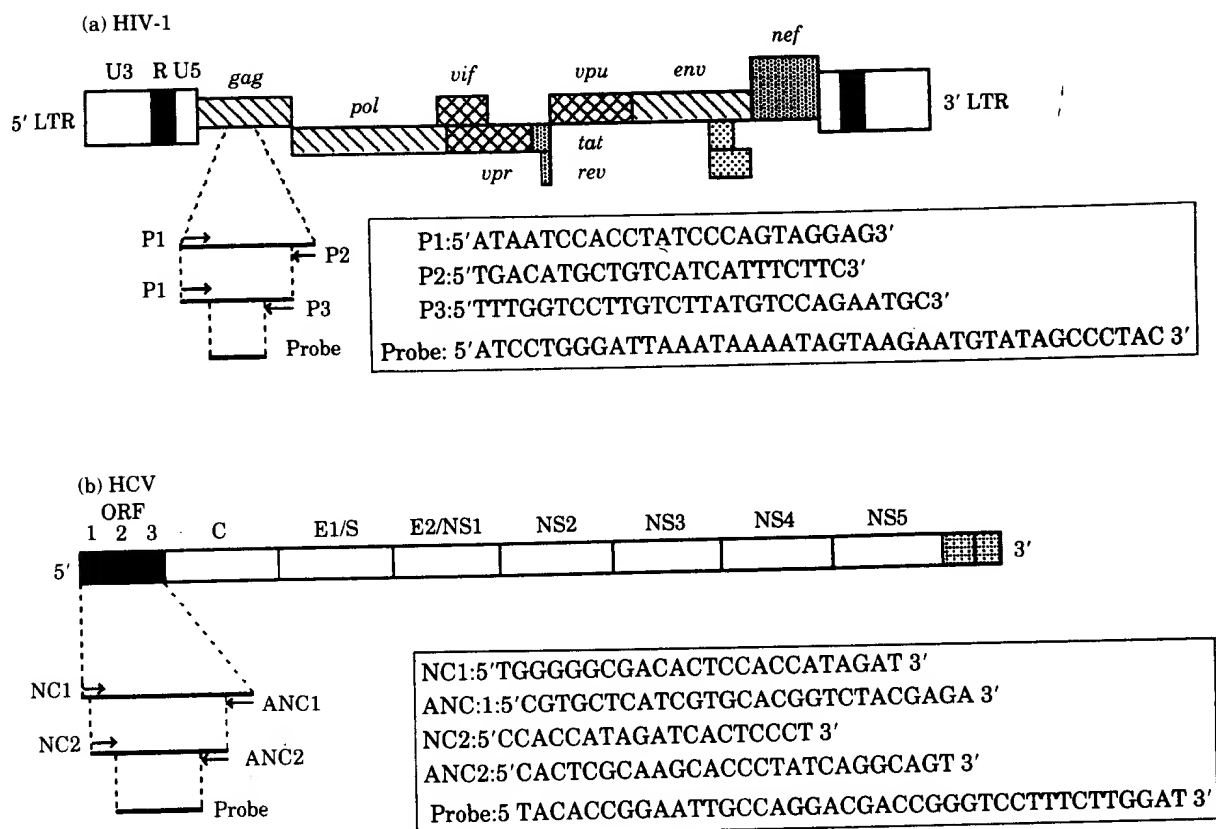


Fig. 1. Positions of the primers and probes used for nested PCR amplification and hybridization of (a) HIV-1 gag and (b) HCV.

The amplification was carried out as described above for HIV-1; 10 µl of the amplified products of both rounds (yielding 340 and 286 bp products, respectively) were subjected to electrophoresis in a 3% agarose gel and stained with ethidium bromide. The presence of a 286 bp band in the second amplification was considered a positive result.

### Hybridization

For some samples, and in order to check the specificity of the amplified products, the DNAs were transferred to a Hybond N<sup>+</sup> membrane (Amersham). Probes specific to HCV (5' TACACCGGAATTGCCAGGACGACCGGGTCCTTTCTTGGAT 3') and to HIV-1 gag (5' ATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTAC 3') were labelled using a digoxigenin oligonucleotide tailing kit (Boehringer). Hybridization was carried out overnight at 52°C and the reaction

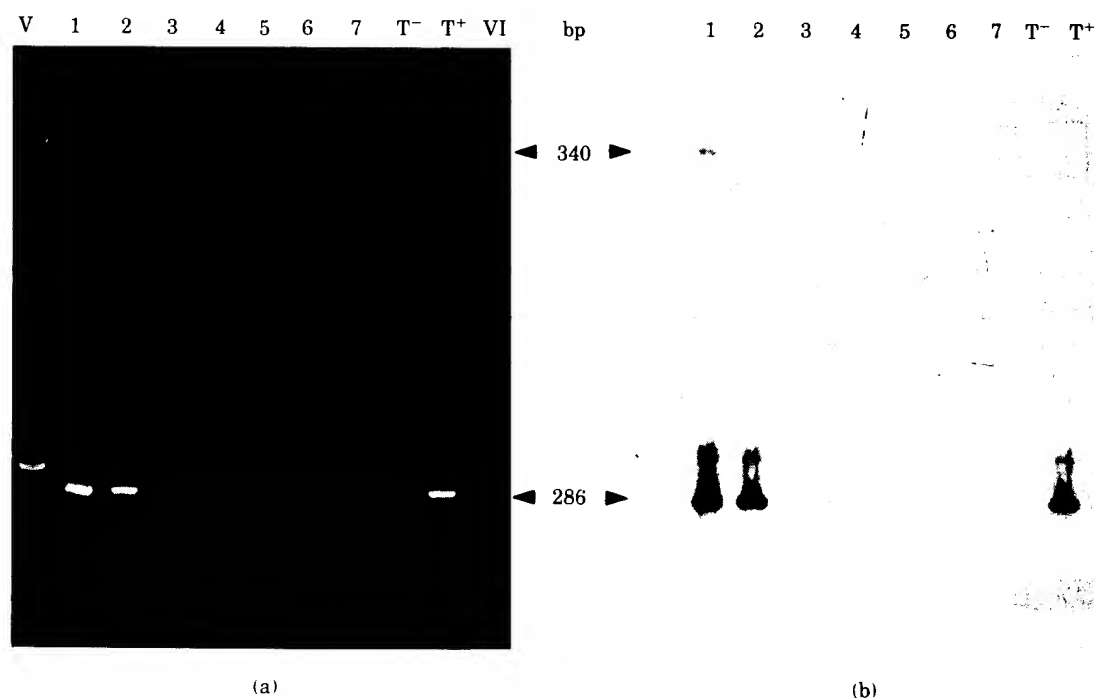
was developed using a digoxigenin luminescent kit (Boehringer).

### Statistical analysis

Chi-square analyses were used to study the repartition of qualitative parameters and Student's *t*-test for comparison of means.

### RESULTS

As described above, the presence of 115 and 286 bp bands in the agarose gel after the second PCR round was considered positive for HIV-1 and HCV RNA, respectively (Fig. 2). The presence of a band after the first round (276 and 340 bp, respectively) was always followed by the appearance of the expected band at the second round. When Southern blot hybridization with probes (specific for HIV-1 and HCV, respectively)



**Fig. 2.** (a) Agarose gel electrophoresis of amplified PCR products from seven anti-HCV reactive sera samples and (b) Southern blot hybridization analysis of HCV amplified products. Upper part, first round of PCR; bottom part, second round of PCR. Lanes V and VI, molecular weight markers; lane T<sup>-</sup> and T<sup>+</sup>, PCR negative and positive controls.

**Table 1.** HCV RNA and hepatitis

Hepatitis	HCV RNA	
	Negative	Positive
Negative	16	5
Positive	6	31

$P < 5 \times 10^{-6}$

was carried out, the specificity of the observed bands could be confirmed in all cases (Fig. 2).

### HCV RNA and hepatitis

Among 58 patients studied with the nested PCR, 36 (62%) exhibited serum HCV RNA; 31 had a chronic hepatitis. The association between chronic hepatitis and circulating HCV RNA was highly significant ( $P < 6 \cdot 10^{-6}$ ) (Table 1). Only six patients with anti-HCV antibodies and chronic hepatitis failed to produce HCV-RNA in plasma; all had another cause of hepatitis: three were treated for tuberculosis, two had systemic CMV infection and one patient received valproic acid.

**Table 2.** HCV RNA and CD<sub>4</sub><sup>+</sup> lymphocytes (at entry)

No. of patients	HCV RNA	CD <sub>4</sub> <sup>+</sup> $\mu\text{l}^{-1}$
36	Positive	199 $\pm$ 180
22	Negative	144 $\pm$ 190

NS

**Table 3.** HCV RNA and  $\beta$ 2 microglobulin

No. of patients	HCV RNA	$\beta$ 2 microglobulin (mg l <sup>-1</sup> )
29	Positive	3998 $\pm$ 1281
22	Negative	3416 $\pm$ 803

NS

### HCV RNA, CD<sub>4</sub><sup>+</sup> lymphocytes and $\beta$ 2 microglobulin

Mean CD<sub>4</sub><sup>+</sup> lymphocytes counts (Table 2) and  $\beta$ 2 microglobulin levels (Table 3) were not significantly different between patients with and without release of HCV-RNA within the serum ( $P = 0.28$  and  $P = 0.12$ , respectively).

**Table 4.** HIV-1 RNA and CD<sub>4</sub><sup>+</sup> lymphocytes

No. of patients	HCV RNA	CD <sub>4</sub> <sup>+</sup> $\mu\text{l}^{-1}$
33	Positive	99 $\pm$ 133
25	Negative	267 $\pm$ 184

P &lt; 0.02

**Table 5.** HIV-1 RNA and  $\beta$ 2 microglobulin

No. of patients	HIV-1 RNA	$\beta$ 2 microglobulin (mg l <sup>-1</sup> )
33	Positive	4974 $\pm$ 2793
25	Negative	3272 $\pm$ 700

P < 4  $\times$  10<sup>-3</sup>**Table 6.** HIV-1 and HCV RNAs

HIV-1 RNA	HCV RNA	
	Negative	Positive
Negative	8	17
Positive	13	20

NS

### HIV-1 RNA CD<sub>4</sub><sup>+</sup> lymphocytes and $\beta$ 2 microglobulin

Overall, HIV-1 RNA was detected in 33 out of 58 (57%) patients. The CD<sub>4</sub><sup>+</sup> lymphocyte counts were lower in patients with circulating HIV-1 RNA ( $P < 1.10^{-5}$ ) (Table 4). Higher  $\beta$ 2 microglobulin levels were observed in patients with positive HIV-1 RNA ( $P < 4 \times 10^{-3}$ ) (Table 5).

### HCV RNA and HIV-1 RNA

Among the 58 investigated patients, eight had negative signals while 20 (34%) had positive signals for both HIV-1 and HCV RNAs; 13 exhibited only serum HIV-1 RNA and 17 only HCV RNA (Table 6). The presence of HCV RNA did not correlate with that of HIV-1 RNA.

## DISCUSSION

The aim of the present study was to determine, in a series of HIV-1 and HCV antibody-positive patients, the percentages of those with active replication of HIV-1, HCV and both. To achieve this goal, we used a nested PCR for detection of HIV-1 and HCV RNAs. We compared the observed data to biological parameters of hepatitis and immunodeficiency.

Thirty-three patients out of 58 (57%) had existence of active replication of HIV-1, as noticed with our nested PCR. These results can be compared to those obtained by other groups ranging from 60% to 100% depending on clinical stages of the patients and CD<sub>4</sub><sup>+</sup> levels.<sup>7-14</sup>

The correlation between the release of HIV-1 RNA and low numbers of CD<sub>4</sub><sup>+</sup> lymphocytes is confirmatory since it is now well recognized that the synthesis of HIV-1, measured by plasma viraemia determination and/or PCR RNA, increases with the decrease of CD<sub>4</sub><sup>+</sup> lymphocytes. Also confirmatory is the fact that low counts of CD<sub>4</sub><sup>+</sup> lymphocytes are associated with an increase of  $\beta$ 2 microglobulin levels.

Thirty-six out of 58 patients (62%) HIV-infected patients exhibited active replication of HCV. Considering that synthesis of HCV RNA is associated with a persistent infection, 62% of HIV-1/HCV co-infected patients in our series are at risk for developing chronic hepatitis.<sup>25</sup> There was no correlation between the presence of HCV-RNA and the immune status (CD<sub>4</sub><sup>+</sup> lymphocytes and  $\beta$ 2 microglobulin) of the patients; therefore, the immunodeficiency does not appear to enhance HCV replication and HCV replication does not appear to result in immune deterioration. Clearly, the active synthesis of HCV is correlated with hepatitis. It is interesting to note that some patients (five in our series) release HCV within the plasma without exhibiting clinical signs of hepatitis, as has been observed by others.<sup>26</sup> In addition, 20 patients out of 58 (34%) exhibited viral replication of both viruses simultaneously; the statistical analysis indicates, however, that there is no correlation between the presence of HIV-1 RNA and that of HCV RNA. Therefore, active replication of HIV-1 and HCV might not affect each other. It has been shown recently<sup>27,28</sup> that HCV can infect PBMC. These new data should stimulate *in vitro* studies of the interactions between HIV-1 and HCV in PBMC cultures, and in the peripheral blood cells of infected patients.

## REFERENCES

1. Coombs R. W., Collier A. C., Allain J. P. et al. (1989). Plasma viremia in human immunodeficiency virus

- infection. *New England Journal of Medicine* **321**, 1626–31.
2. Ho D. D., Moudgil T. & Adam M. (1989). Quantitation of the human immunodeficiency virus type 1 in the blood of infected persons. *New England Journal of Medicine* **321**, 1621–5.
  3. Rouzioux C., Puel J., Agut H. et al. (1992). Comparative assessment of quantitative HIV viremia assays. *AIDS* **6**, 373–7.
  4. Venet A., Lu W., Beldjord K. et al. (1991). Correlation between CD4 cell counts and cellular and plasma viral load in HIV-1 seropositive individuals. *AIDS* **5**, 283–8.
  5. Mathez D., Paul D., De Belikovsky C. et al. (1990). Productive human immunodeficiency virus infection levels correlate with related manifestations in the patient. *Proceedings of the National Academy of Sciences, USA* **87**, 7438–42.
  6. Pan L. Z., Werner A. A. & Levy J. A. (1993). Detection of plasma viremia in human immunodeficiency virus-infected individuals at all clinical stages. *Journal of Clinical Microbiology* **31**, 283–8.
  7. Hart C., Schochetman G., Spira T. et al. (1988). Direct detection of HIV RNA expression in seropositive subjects. *Lancet* **II**, 596–9.
  8. Lewis D. E., Minshall M., Wray N. P. et al. (1990). Confocal microscopic detection of human immunodeficiency virus RNA – producing cells. *Journal of Infectious Diseases* **162**, 1373–8.
  9. Ottman M., Innocenti P., Thenadey M. et al. (1991). The polymerase chain reaction for the detection of HIV-1 genomic RNA in plasma from infected individuals. *Journal of Virological Methods* **31**, 273–84.
  10. Holodniy M., Katzenstein D. A., Sengupta S. et al. (1991). Detection and quantification of human immunodeficiency virus RNA in patient serum by use of the polymerase chain reaction. *Journal of Infectious Diseases* **163**, 862–6.
  11. Semple M., Loveday C., Weller I. et al. (1991). Direct measurement of viraemia in patients infected with HIV-1 and its relationship to disease progression and zidovudine therapy. *Journal of Medical Virology* **45**, 35–8.
  12. Schnittman S. M., Greenhouse J. J., Lane H. C. et al. (1991). Frequent detection of HIV-1 specific mRNAs in infected individuals suggests ongoing active viral expression in all stages of disease. *AIDS Research and Human Retroviruses* **7**, 361–7.
  13. Michael N. L., Vachey M., Burke D. S. et al. (1992). Viral DNA and mRNA expression correlate with the state of HIV-1 infection in humans: evidence for viral replication in all stages of HIV disease. *Journal of Virology* **66**, 310–16.
  14. Piatak M., Saag M. S., Yang L. C. et al. (1993). High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**, 1749–54.
  15. Delord B., Ottman M., Schrive M. H. et al. (1992). Analysis of HIV-1 expression *in vivo* with *in situ* hybridization and the polymerase chain reaction. *Molecular and Cellular Probes* **6**, 215–21.
  16. Kuo G., Choo Q. L., Alter H. J. et al. (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244**, 362–4.
  17. Alter H. J., Purcell R. H., Shih J. W. et al. (1992). Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *New England Journal of Medicine* **321**, 1494–500.
  18. Weiner A., Kuo G., Bradley D. et al. (1990). Detection of hepatitis C viral sequences in non-A, non-B hepatitis. *Lancet* **I**, 1–3.
  19. Garson J. A., Tedder R. S., Briggs M. et al. (1990). Detection of hepatitis C viral sequences in blood donations by nested polymerase chain reaction and prediction of infectivity. *Lancet* **335**, 1419–22.
  20. Cristiano K., Di Bisceglie A. M., Hoonagle J. H. & Feinstone S. M. (1991). Hepatitis C viral RNA in serum of patients with chronic non-A, non-B hepatitis: detection by the polymerase chain reaction using multiple primer sets. *Hepatology* **14**, 51–5.
  21. Inchauspe G., Abe K., Zebedee S. et al. (1991). Use of conserved sequences from hepatitis C virus for the detection of viral RNA in infected sera by polymerase chain reaction. *Hepatology* **14**, 595–600.
  22. Nedjar S., Biswas R. M. & Hewlett I. K. (1991). Co-amplification of specific sequences of HCV and HIV-1 genomes by using the polymerase chain reaction assay: a potential tool for the simultaneous detection of HCV and HIV-1. *Journal of Virological Methods* **35**, 297–304.
  23. Okamoto H., Okada S., Sugiyama Y. et al. (1990). The 5'-terminal sequence of the hepatitis C virus genome. *Japanese Journal of Experimental Medicine* **60**, 167–77.
  24. Han J. H., Shyamala V., Richman K. H. et al. (1991). Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5' untranslated region and poly (A) tails at the 3' end. *Proceedings of the National Academy of Sciences, USA* **88**, 1711–15.
  25. Alter M. J., Margolis H. S., Krawczynski K. et al. (1992). The natural history of community-acquired hepatitis C in the United States. *New England Journal of Medicine* **327**, 1899–905.
  26. Alberti A., Morsica G., Chemello L. et al. (1992). Hepatitis C viraemia and liver disease in symptom-free individuals with anti-HCV. *Lancet* **340**, 697–8.
  27. Signego A. L., Macchia D., Monti M. et al. (1992). Infection of peripheral mononuclear blood cells by hepatitis C virus. *Journal of Hepatology* **15**, 382–6.
  28. Shimizu Y. K., Iwamoto A., Hijikata M. & Purcell R. H. (1992). Evidence for *in vitro* replication of hepatitis C virus genome in a human T-cell line. *Proceedings of the National Academy of Sciences, USA* **89**, 5477–81.